УДК 004.89:004.93

М.С. Клименко
Інститут проблем штучного інтелекту МОН і НАН України, Україна
пр. Академіка Глушкова, 40, м. Київ, 03680

УДОСКОНАЛЕНИЙ МЕТОД РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ ДИКТОРА ІЗ СЕМАНТИЧНИМ АНАЛІЗОМ ЗМІСТУ

M.S. Klymenko
Institute of artificial intelligence problems of MES and NAS of Ukraine, Ukraine
40, Academician Hlushkov av., Kyiv, 03680

IMPROVED METHOD OF EMOTIONAL CONDITION RECOGNIZING BY VOICE USING SEMANTIC ANALYSIS OF CONTENT

У статті запропоновано удосконалення методу розпізнавання емоційного стану людини за голосом шляхом розширення множини ознак емоцій. Додано ознаку емоційного акценту висловлювання на основі семантичного аналізу. Описано вимоги до модифікації структури системи розпізнавання, зображено схему бази онтології та термінів. Проведено числове дослідження, яке показало підвищення ймовірності розпізнавання емоцій емоційних станів порівняно із результатами без використання семантичного аналізу.

Ключові слова: онтологія, акустичні характеристики емоцій, концепт, модель сумішей Гауса

In the article the improved method of emotional condition recognizing by voice is described. The improvement was made by feature vector expansion. The feature of emotional tone of the statement based on semantic analysis is added. Requirements for structure modification the recognition system and the schema of ontology and term database are described. A numerical research showed an increase of emotional recognition probability compared to the results without the use of semantic analysis.

Keywords: ontology, acoustic characteristics of emotions, concept, the Gaussian mixture model

Вступ

Задача розпізнавання емоційних станів людини за її голосом на сьогодні не є повністю вирішеною через високу варіативність та індивідуальність проявів емоцій, що значно знижує ефективність існуючих методів.

На відміну від зміни артеріального тиску або частоти скорочення серця, характеристики прояву емоцій за голосом є менш помітними, але становлять більший інтерес у сфері практичного застосування в медицині та засобах безпеки через можливість безконтактної роботи. А використання засобів аудіозапису вигідно спрощує реалізацію даних систем у порівнянні із методами розпізнавання низки емоцій за зображеннями міміки людини.

Дана робота ставить за мету удосконалити метод, запропонований у [1]. За допомогою методу із використанням фразових моделей вдалося досягти рівня розпізнавання 84% ± 5,3% ($p < 0,05$) проявів емоцій на множині 20 дикторів. Значну кількість помилок розпізнавання зафіксовано через «невпевненість» класифікатора у своєму рішенні. Таким чином, у даній статті запропоновано шлях до підвищення робастності методу та збільшення розділової відстані між моделями емоційних проявів у визначеному ознаковому просторі.

Постановка задачі

Із мети даної роботи випливають наступні задачі:

1. Виконати аналіз недоліків існуючого методу розпізнавання емоційного стану за голосом.

2. Розробити та описати удосконалений метод розпізнавання проявів емоцій.
3. Провести чисельне дослідження ефективності удосконаленого методу.

Аналіз наявних результатів класифікації проявів емоцій

Для моделювання голосових проявів обрано 7 емоцій з переліку К. Ізарда [2]: інтерес (*E1*), радість (*E2*), страждання (*E3*), відраза (*E4*), страх (*E5*), гнів (*E6*), сором (*E7*). Ці емоції обрані для спрощення реалізації розпізнавання у даній роботі, оскільки вони є достатньо різними як за семантикою, так і за проявом.

Однак була помічена згрупованість помилок розпізнавання між певними моделями, що свідчить про недостатню розділову здатність ознак або низьку якість навчання моделей для врахування особливостей проявів даних емоцій. Особливо наочною дана згрупованість помилок простежується на відповідних графіках попарного порівняння ймовірності приналежності тестових зразків до моделей емоційних проявів (рисунок 1), де помилковою класифікацією буде та, якій відповідає $P_2(E_x) > P_1(E_x)$.

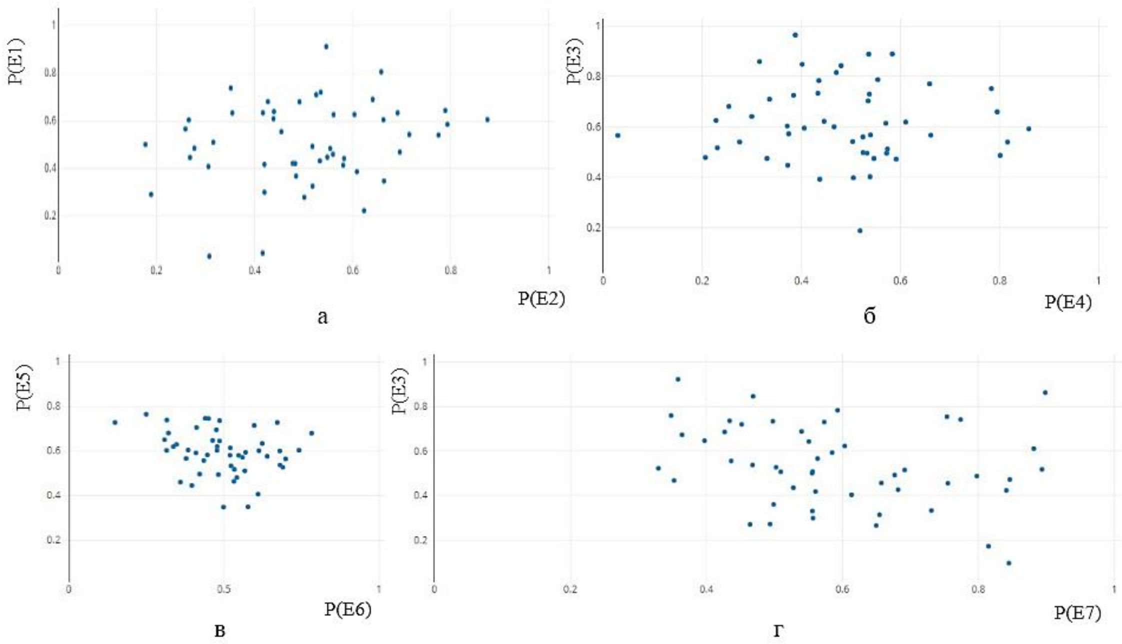


Рис. 1. Графіки попарного порівняння ймовірності приналежності тестових зразків до моделей емоційних проявів: *а* – інтересу та радості, *б* – страждання та відрази, *в* – страху та гніву, *г* – страждання та сорому.

Слід зазначити, ймовірності приналежності тестових зразків до моделей інших емоційних проявів були обчислені значно нижчими, тому у своїй більшості не впливають на результат розпізнавання. Розглянемо більш детально причини такої поведінки класифікатора.

Для моделювання особливостей прояву емоційних станів було обрано наступні параметри:

1. Нормовані значення енергетичного спектра, відносний час перебування сигналу у смугах енергетичного спектра та відносна потужність спектра мовлення у смугах.
2. Значення компоненту гістограми розподілу частоти основного тону (ЧОТ).
3. Кепстральні коефіцієнти.

4. Просодичні характеристики. Основні параметри, які необхідні для визначення типу інтонаційної конструкції:
   – початкова та кінцева ЧОТ;
   – максимальна та мінімальна ЧОТ та час цих значень (у межах контуру);
   – середня частота (усереднене значення ЧОТ у межах контуру);
   – час половини частоти (позиція значення середньої ЧОТ у відсотках від довжини усього фрагмента);
   – швидкість зміни тону (середня швидкість зростання чи спаду тону на відрізку, Гц/мс).
5. Екстралінгвістичні події (наявність у фрагментах пауз мовленнєвого сигналу кашлю, зітхань, плачу, сміху).

Виходячи із задачі, недоцільно використовувати ознаки, що значною мірою залежать від особливостей диктора. Тому параметри ЧОТ у моделях зазначені не абсолютно (в Гц), а відносно (% від усередненого значення). Таким чином, у методі застосована низка ознак різного роду для максимального охоплення характеристик проявів емоцій у голосі, а також частково нівельований вплив індивідуальності дикторів на формування моделей емоцій [3, 4].

Незважаючи на чималий перелік ознак, окремо їх розділова здатність для врахування особливостей проявів певних груп емоцій є вкрай низькою. Наприклад, прояв інтересу у голосі помітно за зростаючим інтонаційним контуром речень та зменшенням пауз між словами до 12%. А у зразках проявів радості швидкість вимови має зростання до 5% і також наявне зростання інтонаційного контуру. Виходячи з цього, моделі, які створені на основі даних характеристик, мають велику зону перетину в однаковому просторі, що спричинює похибку при обчисленні ймовірності приналежності зразків доданих моделей. Аналогічний перетин значень ознак помітний і в інших парах схожих емоційних проявів.

Для даної задачі використано метод сумішей Гауса, який не спирається на специфіку параметрів і широко зарекомендував себе з векторами ознак великого порядку, зокрема у сфері розпізнавання звукових образів [5]. Модель у просторі ознак представляється у вигляді багатовимірного ймовірнісного розподілу та описується векторами математичного очікування, коваріаційною матрицею і ваговими коефіцієнтами сумішей кожного компонента.

Отже, проаналізувавши чинники помилок класифікації проявів емоцій, видно, що метод потребує удосконалення насамперед у підвищенні розділової здатності ознак. Оскільки набір ознак був визначений як оптимальний із множини акустичних просодичних та екстралінгвістичних ознак, то набір пропонується розширити ознаками іншого роду, які досі не були враховані. Саме такою ознакою є семантична складова фрагмента мовлення. Застосування семантичного аналізу гіпотетично має не тільки підвищити ймовірність розпізнавання близьких за іншими ознаками емоційних проявів, але й дозволить розпізнавати інші тональності емоцій, які визначаються виключно за сумісною оцінкою просодики та семантики (наприклад, гумор та сарказм). У даній роботі зупинимось на визначеній вище множині проявів 7 емоцій для порівняльного дослідження із попередньою роботою [1].

Опис удосконаленого методу

У першу чергу необхідно внести зміни до схеми системи розпізнавання емоцій за голосом [6]. Оновлена структурна схема представлена на рисунку 2. До схеми додано базу термінів, за сукупністю яких можливо визначити емоційний акцент висловлювання.

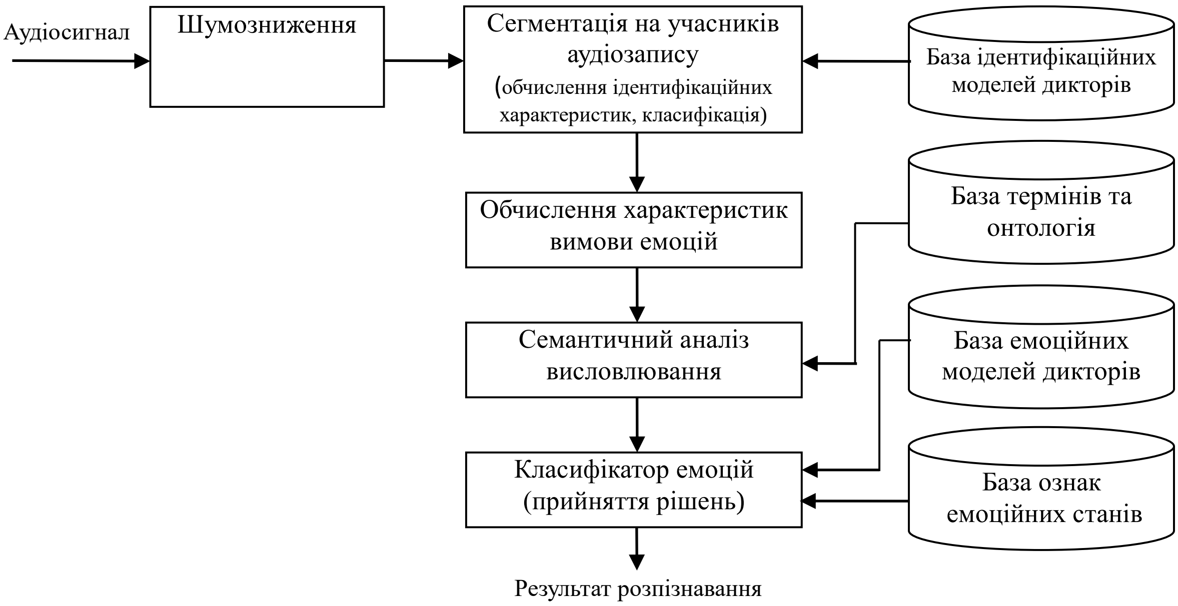


Рис. 2. Структурна схема запропонованої системи розпізнавання емоцій за голосом

Терміни у базі мають бути поєднані між собою у певну онтологію за емоційними ознаками, що визначаються у процесі навчання по корпусу текстів. Також до схеми додано блок семантичного аналізу, який передує безпосередній класифікації проявів емоцій. Результат семантичного аналізу – визначення онтології, що є найближчою до заданого висловлювання.

Важливим є абстрагування текстових термінів від концепцій усередині онтології (рисунок 3). Це необхідно для вирішення проблеми синонімів у межах мови та розширення концептів термінологією інших мов.

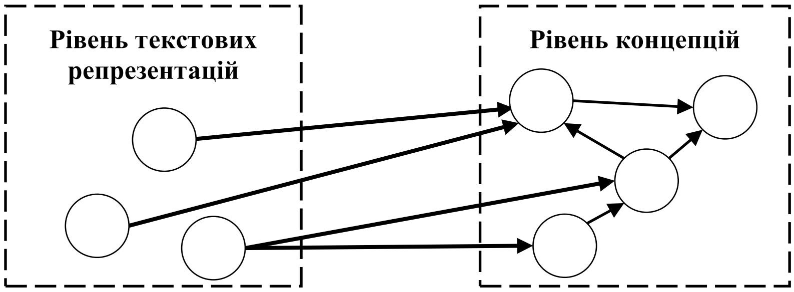


Рис. 3. Схема абстрагування текстових термінів від концепцій усередині онтології

Враховуючи наведені вище вимоги до системи розпізнавання емоцій, запропоновано наступну структуру бази онтології (рисунок 4). У квадратних дужках зазначені необов'язкові поля.

Запропонована структура дозволяє визначати приналежність висловлювання до прояву емоцій не тільки за певними концептами (або термінами, що їх репрезентують), а й за взаємозв'язком даних концептів, тобто множиною атрибутів між ними. Такий підхід повинен підвищити розділову здатність семантичного аналізу емоційних проявів.

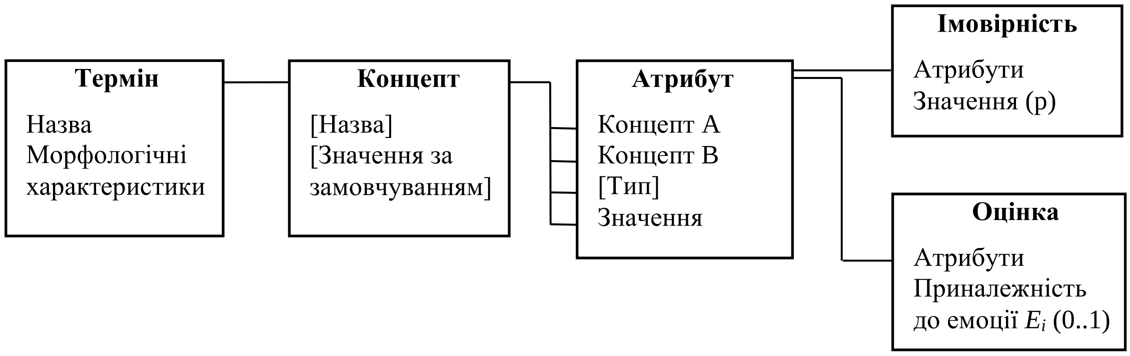


Рис. 4. Структура бази онтології системи розпізнавання емоцій

Визначення семантичного зв'язку *T* між концепом K_T та онтологією $K_x \in O$ у даному випадку набуває вигляду:

$$T = \min_{1 \le i \le n\, i <> T} D(K_T, K_x),$$

де n – розмір онтології K_x.

Числове дослідження ефективності методу

Числове дослідження виконувалось на базі звукозаписів імітації проявів 7 емоцій 20 дикторами віком від 22 до 38 років, сформованих за умов, зазначених у [1]. Автоматизоване розпізнавання тексту виконувалось за допомогою загальнодоступного сервісу Google Cloud Speech-to-Text із ручною корекцією тексту для підвищення якості оцінки запропонованого удосконалення, оскільки розпізнавання тексту не є метою даної роботи.

Результати числового дослідження показали рівень правильного розпізнавання 89% ± 6% ($p < 0,05$) проявів емоцій на множині 20 дикторів. Серед 4 підмножин емоцій, у яких спостерігалась згрупованість помилок розпізнавання, рівень помилок вдалося знизити на 9% ± 4% ($p < 0,05$).

Простежується відносна згрупованість помилок результатів, проте її амплітуда значно зменшилась, що може свідчити про доцільність використання семантичного аналізу як удосконалення методу розпізнавання проявів емоцій.

Висновки

У статті запропоновано удосконалення методу розпізнавання емоцій за голосом за рахунок розширення множини ознак емоцій. Семантична ознака тону висловлювання дозволила знизити помилку розпізнавання груп близьких за ознаками емоцій. Продовженням даної роботи може стати розширення множини емоцій, дослідження робастності удосконаленого методу на великій тестовій множині, а також реалізація повної системи розпізнавання емоцій за голосом із використанням локального методу розпізнавання тексту.

Література

1. Метод розпізнавання емоційного стану диктора за фразовими моделями / Клименко М.С. // Штучний інтелект. – 2017. – №2. – С. 52-59.
2. Изард К.Э. Психология эмоций. - СПб: Издательство "Питер", 1999. - 464 с.
3. Alpert M. Reflections of depression in acoustic measures of the patient's speech / M. Alpert, E.R. Pouget, R.R. Silva // Journal of Affective Disorders. – 2001. – №66. – P. 59–69.
4. Banziger T. The role of intonation in emotional expressions / T. Banziger, K.R. Scherer // Speech Communication. – 2005. – №46. – P. 252–267.

5. Vydana H.K. Improved emotion recognition using GMM-UBMs / H.K. Vydana, P.P. Kumar, K.S.R. Krishna, A.K. Vuppala // 2015 International Conference on Signal Processing and Communication Engineering Systems, Guntur. – 2015. – P. 53-57.
6. Клименко М.С. Розробка структури системи розпізнавання емоційного стану диктора / М.С. Клименко, Ф.В. Фомін // Штучний інтелект. – 2016. – № 1. – С. 17-26.

Literatura

1. Metod rozpiznavannia emotsiinoho stanu dyktora za frazovymy modeliamy / M.S. Klymenko // Shtuchnyi intelekt. – 2017. – №2. – S. 52-59.
2. Izard K.E. Psikhologiya emotsiy. - SPb: Izdatelstvo "Piter". 1999. - 464 s.
3. Alpert M. Reflections of depression in acoustic measures of the patient’s speech / M. Alpert, E.R. Pouget, R.R. Silva // Journal of Affective Disorders. – 2001. – №66. – P. 59–69.
4. Banziger T. The role of intonation in emotional expressions / T. Banziger, K.R. Scherer // Speech Communication. – 2005. – №46. – P. 252–267.
5. Vydana H.K. Improved emotion recognition using GMM-UBMs / H.K. Vydana, P.P. Kumar, K.S.R. Krishna, A.K. Vuppala // 2015 International Conference on Signal Processing and Communication Engineering Systems, Guntur. – 2015. – P. 53-57.
6. Klymenko M.S. Rozrobka struktury systemy rozpiznavannia emotsiinoho stanu dyktora / M.S. Klymenko, F.V. Fomin // Shtuchnyi intelekt. – 2016. – № 1. – S. 17-26.

RESUME

M.S. Klymenko

Improved method of emotional condition recognizing by voice using semantic analysis of content

The article describes the improvement of the method for emotional condition recognizing by voice. The analysis of weaknesses of existing method for emotional condition recognizing by voice is performed. The analysis showed that the set of acoustic, prosodic and extralinguistic characteristics does not have sufficient separation ability to describe the features of emotions in the created sign space. 4 subsets of emotions have been identified that have a significant cross-sectional feature values.

The feature vector is proposed to be expanded by feature of another kind that have not yet been taken into account. It is the semantic component of the speech fragment. The use of semantic analysis hypothetically should not only increase the likelihood of the recognition of close emotional manifestations. It will also allow to recognize other emotional tones that are determined solely by a consistent assessment of prosodic and semantics (for example, humor and sarcasm). The requirements for modification of the recognition system structure are described. The schema of ontology and term database is proposed. This structure allows to determine the belonging of the statement to the manifestation of emotions, not only according to certain concepts (or terms represented by them). Such an approach should increase the separation ability of semantic analysis of emotional manifestations.

The numeric research among the fragments of 20 speakers showed an average probability of emotion recognition of 89% ± 6% ($p < 0.05$) emotional manifestations on the set of 20 speakers, which exceeds the similar results of recognition without semantic analysis of speech.

Надійшла до редакції 30.01.2018